DOCUMENT RESUME

ED 052 628 24 EM 009 100

AUTHOR Block, Karen K.; And Others
TITLE Drill-And-Practice in CAI Spelling: Word Ratings and Instructional Treatment. Project Interim Report Number 1.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-5-0253
PUB DATE Apr 71
CONTRACT OEC-4-10-158(010)
NOTE 41p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Auditory Discrimination, *Computer Assisted Instruction, *Predictor Variables, *Spelling, *Spelling Instruction, Teaching Methods

ABSTRACT

A pilot study investigated the role of one response index (specifically, subject's ratings of their spelling accuracy) that was presumed to be predictive of the amount of practice needed to acquire the spelling of a word and gathered data relevant to the nature of practice needed on a word. The study was conducted to aid in the design of a drill-and-practice, computer-assisted instruction (CAI) spelling program that would complement some recently developed tutorial CAI spelling programs and, at the same time, would expand the range of the spelling curriculum taught via CAI. A brief overview of the interface between tutorial CAI spelling and drill-and-practice CAI spelling serves to introduce the study. Elementary school students were asked to rate themselves on how well they thought they could spell a word after hearing the word spoken. They were then asked to spell the word. The students appeared to be able to predict accurately those words that they could not spell correctly, but were not able to predict from auditory input alone the words they could spell correctly. An examination of this data, together with the results of a survey of the spelling strategies used by the subjects, has implications for the design of instructional programs. (JY)

BR-5-0253
PA 24
EM

UNIVERSITY OF PITTSBURGH - LEARNING R&D CENTER

1971/16

DRILL-AND-PRACTICE IN CAI SPELLING: PROJECT INTERIM REPORT # I. WORD RATINGS AND INSTRUCTIONAL TREATMENT

KAREN K. BLOCK, SHIRLEY TUCKER, AND NANCY B. PESKOWITZ

EM 009 108
ED052628

ED052628

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

DRILL-AND-PRACTICE IN CAI SPELLING:

PROJECT INTERIM REPORT #1.

WORD RATINGS AND INSTRUCTIONAL TREATMENT

Karen K. Block

Shirley Tucker and Nancy B. Peskowitz

Learning Research and Development Center

University of Pittsburgh

April, 1971

The research reported herein was supported by the National Science Foundation and by the Learning Research and Development Center supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Science Foundation or of the Office of Education and no official endorsement should be inferred.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to Mr. Roy Creek, Assistant Director of Falk Laboratory School and Mrs. Linda Binstock, teacher of the mid-group children, for their assistance and co-operation in selecting and scheduling the Ss for this research.

ABSTRACT

This report presents preliminary notions regarding the interface between drill-and-practice and tutorial CAI in spelling in terms of what students might be learning and how it relates to external program characteristics. One line of research into the design of drill-and-practice programs is noted and data from a pilot study comparing predicted and subsequent spelling accuracy are discussed. The intended audience includes learning and instructional psychologists and professional educators.

Drill-and-Practice in CAI Spelling:

Project Interim Report #1.

Word Ratings and Instructional Treatment

Karen K. Block

Shirley Tucker and Nancy B. Peskowitz

University of Pittsburgh

Investigations relevant to the design of optimal instructional treatments frequently take the form of empirical searches for those values of variables operating in the instructional situation which provide optimal outcomes. If drill-and-practice is prescribed, then two parameters of instructional design that are of major importance are (1) The amount of practice: Length of time spent practicing, number of practice trials, etc., and (2) The nature of the practice: Practice may be very specific to the nature of the behavior taught; for example, it may consist of repeated practice of the terminal behavior, or of the prerequisites in combination with the terminal behavior. Practice may also be less specific to the behavior taught. It may require discriminations and decisions that are prerequisites for transfer and/or retention of the behaviors taught and, as such, may take place in a broader context than the original instructional environment. In order to make decisions about the values of these two parameters, some information about the acquisition stage of the processes to be learned must be gathered. The validity of this information must then be established through empirical

investigation. The present pilot study reports the results of a preliminary investigation of one response index (specifically, subjects' ratings of their spelling accuracy) that was presumed to be predictive of the amount of practice needed to acquire the spelling of a word. The study also represents a preliminary attempt to gather data relevant to the nature of practice needed on a word.

The nature of the pilot study was influenced by a number of considerations. It was planned that the results of the study would be relevant to the design of a drill-and-practice CAI spelling program that would complement some recently developed tutorial spelling programs and, at the same time, would expand the range of the spelling curriculum taught via CAI. Although the interface between tutorial spelling and drill-and-practice spelling will be elaborated upon elsewhere (Block and Butler, in preparation), some brief statement of the major arguments here will serve to clarify the design of the study and communicate some recently developed ideas about CAI.

Tutorial CAI

The tutorial instructional component of CAI spelling consists primarily of instruction designed to be isomorphic to the performance of good spellers. This tutorial instruction is designed by analyzing and identifying the components of competent spelling performance. There are three major components of performance that can be identified in the behavior of skilled spellers. The first is accurate auditory analysis: the capability of Ss to analyze a complex speech signal into

the phonemes that comprise the language. In addition to phoneme identification, auditory analysis prerequisite to good spelling performance requires Ss to locate the position of target phonemes in syllables (as the initial, medial, or terminal sound), to be able to isolate and identify adjacent phonemes (as in segmentation) and to discriminate stressed from unstressed syllables.

The second major component of skilled spelling performance is the successful application of sound-to-letter spelling rules. This component requires accurate auditory analysis which provides the cues that Ss can use to choose the correct graphemic representation of a sound. For example, the /k/ in cat can be represented graphemically by c, k, ch, ck. When cues such as the position of /k/ in the syllable, adjacent phonemes and graphemes, etc., are considered, each of the options becomes differentially appropriate. For example, c and k are the most frequent options for the initial position in a syllable; k is used when i or e is the succeeding grapheme; c is used for other succeeding graphemes. Successful rule application in the case of optional sound-to-letter correspondences requires auditory analysis skills which provide the target sound, a knowledge of the tenable options for that sound, and a knowledge of tenable values of the cues as they determine the appropriate spelling option as a function of context.

The third major component of competent spelling performance is a knowledge of the morphology of the language. This allows Ss to discriminate "root" words from suffixed and prefixed forms in order that structural rules, such as the "ing" rule, may be applied to generate the

spelling of these derived forms. Successful application of structural rules also depends upon accurate auditory analyses, the outcome of which provides cues for correct rule application. For example, in the "ly" rule, the terminal y is changed to i if the terminal sound is /e/ (as in happy - happily), and not changed if the terminal y is sounded /i/ (for example, shy - shyly).

To illustrate the relation between competent spelling performance and our tutorial instruction, some modules from the tutorial CAI spelling program OPTIONS will serve well. In OPTIONS, we have assumed Ss are competent in auditory analysis; the objective of the program is to teach mappings of the form: one phoneme to many graphemes. The introductory module, OBSERVE, acquaints Ss with the range of graphemic variability (for example, k and c for initial /k/) by presenting words that are spelled using these options. Module CUE requires S to identify those environmental cues surrounding the /k/ sound which are relevant to the choice of spelling option (for example, succeeding phonemes). Modules SORT and TABLE require S both to identify the cues and then to use these cues in the choice and/or construction of the appropriate graphemic option. The nature of the tutorial instruction is such that it is isomorphic to those components that we have identified in the behavior of good spellers.

The broadest objective for spelling instruction is to insure that Ss acquire spelling rules that have high utility. These rules should be useful in the sense that they have predictive validity which results

in above chance spelling accuracy of words or parts of words. The rules should also be chosen so that they can account for a large portion of the terminal objectives of a spelling curriculum. The objective of the tutorial spelling program is to teach useful rules through teaching the students to analyze words on dimensions that are salient to the development of efficient spelling rules; or even to teach the rules themselves. Since rule learning in spelling requires the discrimination of similarities and differences among words (for example, whether several words contain the same target phoneme, and whether this phoneme occurs in the same or different syllabic positions), then the input to tutorial instruction must necessarily be sets of words that are similar along the target dimensions[1] and less similar along other dimensions less relevant to the strategy. Thus, tutorial instruction requires a specifiable and controlled input list of words.

Drill-and-Practice CAI

Within CAI spelling, drill-and-practice is an appropriate instructional treatment when the student's task is to learn and retain the spelling of a list of words that are much more heterogeneous than those that are treated in tutorial spelling. Words may be designated as

---

[1]At the present time, similarities among words have been restricted to logically and externally defined similarities such as phonetic likenesses and structural similarities. It may be the case that other dimensions of similarity exist that influence the mode of generating a spelling response; these must await empirical (or theoretical) discovery.

unrelated, or heterogeneous, and treated in drill-and-practice because they have no obvious external similarity (phonetic or structural) or because other similarity dimensions that would form the basis for tutorial instruction are indeterminate (they have not been discovered and validated). Additionally, it is the case that most words rather than differing on one or two critical dimensions (such as the simple monosyllabic words currently treated in tutorial spelling) differ on a great many dimensions, and require the joint use of several rules taught in tutorial CAI spelling for the production of a correct spelling. When the objective of instruction is to provide S with practice on several rules concurrently, then drill-and-practice is a reasonable prescription.

In addition to differences in input word lists, drill-and-practice and tutorial instructional strategies differ in terms of how they correspond to what the student learns. Instruction is identified as tutorial instruction when it is explicitly derived from an analysis of competent performance and it forms an isomorphism with that performance. It is assumed that the components of competent performance are the lower bounds to what is learned, i.e., the student at the very least learns the information he is taught (given that he reaches criterion and the instructional situation does not allow him to reach criterion unless he has learned some set of the instructional objective); and he may even learn more.

An illustration of this point can be found in the Trabasso and Bower (1968) studies in attention. These studies were run in situations

with relevant and redundant cues. Some Ss learned one or the other cue (if either was learned, S could reach criterion) and some Ss learned both. With carefully constructed tutorial instruction having acquisition criteria that are "tight," one is able to infer with some confidence, that the student learned what he was supposed to learn, and that it is isomorphic to the components of instruction and competent performance. When what is learned is tested, it is possible to make predictions regarding the kind of items that will be passed, and failed, on the basis of the knowledge of what was learned.

With extant drill-and-practice instructional treatments, it is much more difficult to arrive at some strong rational identity between the instructional treatment and the processes or structures that are acquired due to that treatment. The treatments themselves give very little information as to what is learned, since the instructional strategy is relatively simple: branching is minimal and based on relatively simple analyses of the response (for example, whether it is right or wrong). The items to which the student responds are less carefully sequenced than in tutorial instruction. Thus, from the nature of the treatment, it is difficult to make extensive analyses and hypotheses regarding what is learned. However, the literature on organization in memory suggests that Ss are acquiring rather sophisticated cognitive structures through their efforts to organize, code, and store the information for later retrieval. Thus, the correspondence between the extant instructional strategy and the components of what is learned is

much less well defined with drill-and-practice than with tutorial instruction. Instructional designers operate on the basis of the assumption that any lack of isomorphism to what should be acquired for competent performance in recall should be compensated for by repetition.[2]

To what extent instruction must be isomorphic to what is learned is a difficult question; its resolution is not yet clear and will depend on the efficiency and success of alternatives to drill. It is quite possible that practice of every word in the context of the rules which generate its spelling would provide needless redundancy to tutorial programs and at the same time inhibit generalization of spelling strategies to novel lists of words and to the demands on spelling behavior in the real world. At the same time, if one does not have firm knowledge regarding the cognitive structures acquired during drill on unrelated words and one must make best guesses, one runs the risk that the nature of practice might be isomorphic to the wrong structures or to diametrically opposed structures, or, that it might inhibit the development of invariant structures. The structures acquired, then, might greatly inhibit the development of efficient strategies (see the literature on encoding specificity and organization, for example, Bower and Winzenz [1969], for support for this argument). With primary consideration given to the fact that the processes learned and retained in spelling drill are not clear from currently available data,

---

[2]This is not an unreasonable assumption as repetition does lead to the acquisition and retention of selected terminal behaviors in a vast variety of learning situations.

and that there is a need to teach unrelated words, it was decided that drill-and-practice currently has a place in spelling instruction. Thus, the investigation of the aforementioned variables will contribute some information to the design of this instruction.

The objective of the CAI drill in unrelated words is to provide the subject with instructional treatments which are sufficient for the learning and retention of the word list. In addition to providing sufficient practice, another criterion for the instructional program is that it be efficient, that is, that it provide more practice on words that are in a lower acquisition stage and less practice on words that are nearly learned. Also, it may be the case that the nature of optimally designed practice may be a function of the stages through which a word must pass and the strategies S must learn to exit a stage that are primary to production of the correct spelling.[3] Thus, if it

---

[3]In the absence of any well developed and validated theoretical notions regarding the processes that drill teaches, we could revert to use of the simplest theoretical construct used to explain acquisition in verbal learning: the notion that items (words to be learned) acquire associative strength across repeated presentations. An item is recalled because its associative strength exceeds some criterial level (threshold) necessary for recall. Because this notion does not have sufficient explanatory power when Ss must learn complex responses (as Restle [1964], for example, has shown in paired-associates learning), we suggest that spelling acquisition consists of multiple stages (in the sense of Markovian learning stages) during which a spelling is differentially available for recall. These stages depend upon characteristics of the word list: the familiarity of the words to be spelled, the extent to which they are generated by reliance on a few in contrast to many spelling rules, etc. Although we will not fully characterize these stages in the present paper, they will be characterized in a forthcoming paper. For the moment "stages" will remain a useful notion for the generation of hypotheses regarding variables that might influence acquisition in drill-and-practice spelling.

can be determined which stages and strategies are necessary to produce a word, then some decisions can be made regarding the amount of practice required and the appropriate characteristics of this practice.

In regard to assessing how near a word is to mastery, the literature of recognition and recall suggests that subjects' confidence ratings provide useful information. Specifically, in applications of Signal Detection Theory to analyses of memory, confidence ratings have been shown to be strongly related to the probability of the correct recall of the second item of an item pair in cued recall, and also strongly related to the latency of response at cued recall (Murdock, 1966). In addition to these empirical relations, the relation of confidence ratings to item trace strength has the status of an assumption in the application of the Wickelgren and Norman (1966) models to rating data.

Signal Detection Analyses have been applied to memory data in attempts to improve on traditional strength measures by separating item trace strength from response biases. Bernbach (1967) and others (Banks, 1970; Donaldson & Glathe, 1970) have argued that in recall tasks (or Type 2 analyses) confidence ratings reflect more than the probability of correct recall; they are additionally a function of the subject's response criterion and also the asymptotic discriminability of the items to be recalled (Bernbach, 1967). For our purposes herein, we are less interested in a confidence rating as a pure measure of recall probability rather we are interested in some response measure (the rating) which is monotonically related to recall probability (in our case the probability of spelling the target word correctly). Thus,

although our ratings may be influenced by the above mentioned factors, they may still serve to differentiate levels of item strength on an ordinal scale and are appropriate measures to investigate in this problem context.

The relationship between accuracy judgments and correct and incorrect spellings is a problem which has been investigated in the spelling literature. Tidyman (1924) tested fourth through eighth grade students for accurate spelling judgments. In the context of a sentence dictation test, Ss were asked to spell each word. After spelling all the words, they were asked to mark each word they spelled as Right, Wrong, or Doubtful. When the papers were corrected and scored, Tidyman reported that, of the 8803 words spelled correctly, 8569 were judged Right (97%), 27 were judged Wrong (.3%), and 207 (2.7%) were judged Doubtful. Thus, Ss were able to judge correctly spelled words very accurately. Of the 1764 words spelled incorrectly, 675 were judged Right (38%), 545 were judged Wrong (31%), and 544 were judged Doubtful (31%). To report the data another way, for words judged as Wrong, nearly all were wrong (95%); for words judged as Doubtful, 72% were wrong; words judged as Right were nearly always correct (93%). These data are encouraging, since they indicate that Ss' accuracy judgments are differentially related to their spelling accuracy. In the Tidyman task, Ss spelled approximately 100 words after which they re-scanned the spellings and rated them (no re-pronunciation was introduced). Presumably their judgments were a function of the difficulty they experienced in attempting to spell the word and any information they gained from scanning and/or reading the graphic stimulus.

Hendrickson and Pechstein (1926) studied ratings of Right, Wrong, and Doubtful by college students of words presumably in their reading vocabulary. Ss were given a sentence dictation test of 50 words. No information regarding the rating task was given until the spelling task was completed. After spelling, they were asked to read their papers carefully and rate words with respect to spelling accuracy. No re-pronunciation occurred.

Hendrickson and Pechstein's data showed that of the words spelled correctly, 84.7% were judged as Right, .8% were judged as Doubtful, and 14.5% were judged as Wrong. Their Ss, then, were fairly accurate judges of correctly spelled words. For the words spelled incorrectly, 52% were judged as Right, 8.2% were Doubtfuls, and 39.9% were judged as Wrong. These results are in general agreement with Tidyman's (namely that incorrect spellings are less likely to be detected), except that college students used the Doubtful category much less frequently.

The investigators found that judgments of Wrong did not reliably predict spelling accuracy, since these words were equally likely to be spelled incorrectly (50.7%) or correctly (49.3%). Doubtful judgments were more likely to be spelled incorrectly (79.6%) and Right judgments indicated a word that was most likely to be spelled correctly (81.3%). The major discrepancy between the results of Tidyman and those of Hendrickson and Pechstein is the greater accuracy of the Wrong judgments in Tidyman. Even when Hendrickson and Pechstein's Doubtfuls (which are more accurate predictors of subsequently incorrect words than are Wrong judgments) are collapsed to the Wrong category, the predictive accuracy

increases only about 10% (from 50.7% to 61%). These investigators also found rather large individual differences in accuracy percentages (14% to 92%) and additionally they report a rather high (.68) correlation with spelling ability (percent correct spellings on the 50 word test).

Hendrickson and Pechstein concluded that college students' spelling consciousness (awareness of the accuracy of their spelling) is generally lower than that of elementary school students. There are some procedural variants that may account for this fact. For one, neither Tidyman nor Hendrickson and Pechstein report the instructions given in the use of the rating scale. For another difference, Tidyman's words were much easier for his Ss than Hendrickson and Pechstein's were for theirs, a factor which presumably might influence Ss' decision criteria regarding rating accuracy. Both sets of Ss had to recycle to read their spellings in order to rate them. It could be the case that elementary school students' misspellings contained the kind of errors that made them more difficult to read (i.e., to produce any recognizable word), and, if their criterion for a correct judgment was that the spelling, when read, produced a recognizable word, then these students would be more likely to detect incorrect spellings and Doubtful spellings. In contrast, college Ss' spelling errors, when read, probably result in readable words. Hence, the readability of an incorrect spelling serves less to aid accuracy judgments and Ss must rely on other factors. In other words, for elementary Ss one would expect Wrong judgments to be very accurate predictors of misspellings (as they

were), since an unreadable word would presumably be misspelled. To give an example, one spelling from our study was "farntare" for furniture. For college Ss, Wrong is a less reliable predictor because possibly both correct and incorrect words are readable. Both sets of data also suggest that Ss judged words as Right when they were very certain that the words were correctly spelled (thus, the high rate of "hits" with correct judgments). As uncertainty about a word increased, judgments of Doubtful and Wrong were made.[4]

The literature on level of aspiration has been concerned with the variables governing the manner in which an individual sets his goal or makes judgments about his expected performance. The data from that literature relevant to this study are twofold: First, the S's prediction of his performance will vary as he is asked to state it in different ways. Diggory (1949) found that the discrepancy between S's last performance and his aspiration level was about twice as great when he was asked to state what he "hoped" to score on the next trial as it was when he was asked what he "expected" to score on the next trial. For our purposes the interest is in judgments of realistic expected performance rather than "hoped" performance; thus, the instructions requested estimates of expected performance. Second, since feedback about performance during an experiment changes the level of

---

[4]No SDT analyses were performed on these data so it is not possible to conclude (as Hendrickson and Pechstein did) that college Ss are less sensitive to (or aware of) their incorrect spellings. This result may have been obtained simply through criterion changes induced by differential a priori ease of spelling or different instructions in the use of the rating scale.

aspiration during the course of that experiment (Lewin, Dembo, Festinger, & Sears, 1944, for example), no feedback regarding the correctness of a spelling was given to the Ss.

This literature has also demonstrated the large effect that procedural variations have on estimates of expected performance and their related validity for predictions of performance (Riccuiti, 1951). Our study included two procedural variants of the rating task to assess the extent of this effect. In one procedure, Procedure I, Ss were asked to indicate their expected spelling performance by choosing one of three rating categories upon hearing each of the 15 words (No, I cannot spell it correctly; Maybe I can; Yes, I definitely can spell it). After rating their expected performance on each word, they were then asked to spell each word on the paper provided for them. The words were presented for spelling in the order they were presented for rating. Thus, they rated all words and they returned to spell the words. After spelling all the words, Ss went through the list a third time, and were asked their strategy for spelling each word. In Procedure II, Ss rated a word upon hearing it and then immediately after rating it, they spelled it. Thus, a word was rated and spelled before a new word was presented. After all words were rated and spelled, Ss were questioned by E about their strategy for spelling each of the words. The major relevant difference between our procedures and Tidyman's is that our Ss were required to rate their spelling performance primarily on the basis of the auditory stimulus alone.

Method

Subjects

The subjects were ten students from the mid-group at Falk School, a laboratory school at the University of Pittsburgh. The children ranged in age from 7 to 9 years and were at the intermediate point in the spelling curriculum which roughly corresponds to grade four. Five Ss were randomly assigned to each procedure; one subject in Procedure I experienced great difficulty with the words selected and could not complete the experiment. One S in Procedure II would not attempt the spelling of one word (chocolate).

Word Materials

The word sample was selected from Hanna, Hanna, and Hodges Power to Spell, Books Four and Five. The words were chosen to be strongly illustrative of at least one of three spelling principles: (1) Words that can be spelled by relying primarily on a sound spelling strategy (the sound is most frequently mapped to only one grapheme). (2) Words that require a decision among graphemic options for a correct spelling. (3) Words that require the application of a structural rule for the correct spelling. The words selected for sound spelling were uniform, pajama, graduate, furniture, spectator; for optional spelling, wrinkle, alley, sponge, chocolate, bruise; and for structural spelling, donkeys, laziness, promotion, shrugging, advisable. All nine subjects spelled and rated all words.

Procedure

The subject was ushered into the experimental room and seated beside the experimenter. The E then read the instructions to S. The instructions explained the purpose of the experiment and asked Ss to evaluate themselves on how well they thought they could spell a word by expressing their judgment in terms of the numbers 1, 2, and 3: 1 means no, I cannot spell the word; 2 means maybe; 3 means yes, I'm sure I can spell it. E pointed out that in front of them was a card that had the rating numbers beside the words yes, maybe, no. Ss were told to write their rating in the column labeled "Ratings" on their data sheet and that either after they had finished rating all the words (Procedure I) or after they had rated each word (Procedure II), they would be asked to write the spelling in the column labeled "Words." Then they were told that in the second part of the study, they would be asked a few questions about how they spelled the words, how they put the letters together, or how they had learned to spell the word. Ss were also told that they could ask to hear a word again, if they wished. E also noted to S the presence of a tape recorder to be "sure I can catch all you say." E then asked for questions; if there were some, the instructions were paraphrased.

Procedure I

A trial began with E pronouncing each word as clearly as possible and as many times as necessary. S then wrote his rating. Then, E pronounced the next word and S rated it; these events continued until the list was completed. Then E told S that he would now spell the words he had rated in the same order that he heard them. The words

were then re-pronounced and S wrote his spelling on the data sheet beside the rating of the same word. The presentation order of the words was randomized for each S.

Procedure II

The major difference between Procedure I and Procedure II was that after rating a word, Ss spelled that word before rating a new word. E re-pronounced the word after it had been rated when she requested the spelling.

For the second phase, Ss were questioned about the manner in which they generated a spelling. During this phase, E paraphrased the general question "Can you tell me how you spelled that word?" and asked for elaborations or clarifications of answers given by S. At the same time, E carefully avoided giving S any cues or hints that might bias his report, or any information regarding the correctness of the S produced spelling. E encouraged a response for every word.

Results and Discussion

Accuracy Analyses

Table I shows the distribution of judgments for all Ss and all words with the two procedures combined. Since both our procedures required Ss to rate on the basis of acoustic information alone, in contrast to Tidyman's reading task, the data from the procedures are combined for comparison to Tidyman. Tidyman's data are included in the table for comparison.

TABLE I

Percentages and Numbers of Judgments

Right (Rated 3), Wrong (Rated 1), and Doubtful (Rated 2)

Passed Upon the Subsequently Correctly Spelled and Incorrectly Spelled

Words for the Two Procedures Combined and for Tidyman's Study.

Spelling

	Pilot Study				Tidyman (1924)			
	Correct		Incorrect		Correct		Incorrect	
	N	%	N	%	N	%	N	%
Rating								
Right	30	57%	19	23%	8569	97.0%	675	38%
Doubtful	18	34%	32	40%	207	2.7%	544	31%
Wrong	5	9%	30	37%	27	.3%	545	31%
Total	53	100%	81	100%	8803	100.0%	1764	100%

One point about these two sets of data must be noted to clarify the domain of succeeding inferences. First, our words were much more difficult to spell than Tidyman's (39% correct vs. Tidyman's 83%). Second, it is not reasonable (arguments reiterated in the introduction) nor possible within the context of these data to determine the extent to which rating accuracy is a separate function of the *a priori* probability of a correct spelling or other procedural variants, and the extent to which it is distinctly influenced by *Ss*' threshold for (or sensitivity to) differences between spellings likely to be correct or to be incorrect. Nor, at the present time is it clear in our minds which task characteristics differentially influence *Ss*' response criteria and/or their sensitivity to the occurrence of the signal (a word spelled correctly). Thus, our interpretations of these data are made primarily in terms of task factors that influence rating accuracy, some of which may influence accuracy through *Ss*' criterion changes or through contributing differentially to threshold changes or through both. The reason for this concern regarding data interpretation arises from data from psychophysical experiments which demonstrate that one task factor present in these comparisons (the *a priori* p [signal]) does influence rating accuracy through influencing the manner in which *Ss* use these rating categories (their response criteria) rather than influencing *Ss*' sensitivity to the differences between the signal and the noise. Thus, any inferences made regarding task factors that influence *Ss*' discriminations of correctly spelled and incorrectly spelled words do not have independent validity. They do, however, provide some interesting

hypotheses for future research and so will be advanced herein, with the caution that the experimental evidence needed to demonstrate their validity is confounded by other factors.

For words subsequently spelled incorrectly, the data in Table I compare quite well to Tidyman's. Ss were more likely to judge a subsequently incorrectly spelled word as Doubtful or Wrong than they were to judge it as Right; the probability of each of these various judgments agrees with Tidyman. The Ss in this study then were not always able to detect a word that would subsequently be spelled incorrectly; 23% of the time subsequently incorrect words were predicted to be Right. For words subsequently spelled correctly, our Ss were much less able to detect these than Tidyman's Ss. The "hit rate" for subsequent corrects was substantially lower for our Ss (57% vs 97%). Thus, our Ss achieved approximately the same hit rate for incorrectly spelled words as Tidyman's Ss, but achieved a much lower detection accuracy for correctly spelled words. In both studies, the detection accuracy for correctly spelled words was higher than for incorrectly spelled words. The suggestion from these data is that detection accuracy for subsequently incorrectly spelled words can be maintained in the absence of the visual cues from the misspelling and the subsequent reading behavior or any other behaviors relevant to judgment decisions that they introduce or permit. In other words, the level of detection accuracy for incorrects can be maintained in the presence

of the recent acoustic information alone.[5] Reading and hearing, then, result in the same distribution of accuracy judgments.

Further analyses of the data lend a bit more support to this statement and suggest one possible interpretation of the manner in which Ss make accuracy judgments on the basis of the auditory signal and the written word. If accuracy decisions must be made on the basis of the auditory stimulus, then Ss might be primarily making judgments of the clarity of their auditory perceptions: the degree of match between what they recall they heard and what they can reproduce. Presumably they might weigh less such factors as the graphemic variability of the sound and knowledge of the structural rules, both of which are additional, but necessary components of spelling accuracy. This implies that incorrectly spelled words rated Wrong should evidence lower acoustic accuracy than those rated Right. When these two sets of words were scored for acoustic errors such as the omission of a pronounced chunk as in the word "farntare"; sound reversals, for example, "shurgging" for shrugging; and letting incorrect graphemic representation of an acoustic chunk count as presence of that chunk, it was found that incorrectly spelled words judged Right had one acoustic deficiency (1/21 = 5%), while those judged Wrong had at least seven (7/31 = 22%).[6] Although these differences are not large, they lend some tentative support to the previous interpretation.

---

[5]It should be recalled that Tidyman's Ss spelled 100 words, then rated them with no re-pronunciation. There probably was minimal accurate acoustic information available about each word when it was rated (assuming words were rated in the order they were spelled).

[6]One rater scored 10 in this group.

If Ss read their misspellings, then words containing acoustic deficiencies are probably less likely to result in recognizable words. Acoustic deficiencies, however, are not the only spelling errors to contribute to reading difficulty; others, such as the choice of an incorrect graphemic option or the misapplication of a structural rule, also contribute. To measure the readability of an incorrect spelling, a scoring system was derived which takes into account various sources of spelling errors and weights these sources by the degree to which they affect readability. These weights were arrived at on the basis of the intuitive judgment of the authors. For acoustic errors, the READ score was tallied 2 for each omitted pronounceable chunk (for example, in "farntare") and tallied 1 for each sounded letter that was omitted (for example, "shugging"). Incorrect graphemic options were scored 2 when they were very low frequency options for a particular position in a syllable (for example, "spictature"), and scored 1 when they were more frequent (for example, "rinkle"). Misapplications of structural rules (when adding "ing," misspellings of a suffix or prefix, forming plurals, etc.) were always scored 1. Although the investigators were unable to get complete agreement on independent scorings, the results were always in the same general direction; what is reported here is the average score from several scorings. For the Right words incorrectly spelled, the mean READ score was 1.22; for the Wrong words, the mean READ score was 1.97. Once again these data, although tenuous, suggest that misspellings rated Wrong are harder to read than those rated Right and that if our Ss had their responses available to read when rating, then the same distribution of accuracy might have

been obtained, predictable through an analysis of readability alone.[7] To demonstrate some independent effect of the influence of viewing the spelling, the subsequently correct spelling data must be reviewed.

The data in Table I also revealed that detection accuracy of subsequently correctly spelled words decreased when only acoustic cues were available. Thus, acoustic cues available from recent pronunciations were not sufficient for the accurate ratings obtained by Tidyman; rating accuracy increased greatly when Ss were permitted to read correct spellings. Presumably when Ss read a correct spelling (and read it correctly), the acoustic characteristics of the word are most likely to be present and are available for continuous re-activation and rehearsal. With these, Ss are then able to focus their attention upon an analysis of the graphemic representation (in terms of a choice of options; option contingencies and a judgment regarding the application of structural rules). Thus, these Ss can spend more time in the second stage of spelling decision making, time which delivers greater accuracy in the detection of correct spellings. For Ss judging on the basis of

---

[7]One point should be clarified: although reading a spelling word results in the production of auditory cues which are used to identify the word and to subsequently judge spelling accuracy, these are not always identical to the E given-S matched auditory cues which were available to our Ss when they predicted subsequent spelling accuracy. Not only are they not necessarily identical, but they may be used in different ways when they are generated in the context of a spelling or are generated through auditory analysis alone. Our argument here is simply that our data (and probably Tidyman's) are loosely consistent with the predictions of a readability analysis, and these predictions match distributions of judgments observed when auditory cues alone are available. There are no data from this study which allow stronger process inferences.

auditory cues alone, these Ss must maintain the auditory cues while they make letter decisions, a task which would presumably interfere with accurate decision making in this second stage. Subsequently correct spellings reveal that Ss did, in fact, make accurate auditory analyses of the words, but either these were not judged to be adequately available for subsequent entry into a letter choice stage, or maintenance of these interfered with accurate decision making in the letter stage, or, possibly, Ss did not have accurate knowledge of optional and structural rules which dictated the spelling of the word. Any of these factors or all of them might have produced the lowered detection rate for subsequent corrects. The data from Table II (to be discussed below) provide some information on these suggestions.

Since it is possible that Ss do not enter the second decision making stage when rating with auditory presentations alone, they may be making spelling accuracy predictions on the basis of their perceptions of the clarity, rehearsability, and recallability of the auditory signal alone. Analysis of the acoustic deficiencies of words spelled incorrectly supports this notion. Additionally, the Protocol Analysis data (to be reported below) revealed that Ss most frequently reported spelling strategy was to "sound out the letter," data which may indicate that Ss spent portions of their time analyzing the auditory stimulus in contrast to making letter choices.

In summary, then, accuracy judgments of subsequently incorrectly spelled words made on the basis of auditory input alone have the same distribution as judgments made on the basis of reading incorrect

spellings. Thus, auditory input was sufficient to maintain rating accuracy (and might even be the critical feature in reading, then rating decisions); the distributions were very much alike because of the strong functional interdependency of reading and auditory analysis. However, the level of rating accuracy was not high. Words predicted "wrong" evidenced more acoustic deficiencies than those predicted "right," lending support to the notion that a judgment of subsequent correct spelling is a function of the degree to which the necessary auditory features of a word-to-be-spelled are (or will be) available.

When correctly spelled words were available to be read, rating accuracy was very high, possibly because Ss were able to spend more time judging the accuracy of letter representations. With an exclusively auditory based rating judgment, Ss were less able to make accurate spelling predictions because of several factors, one among which was the necessity to maintain the auditory characteristics relevant to letter choice while considering (if any consideration at all was made) additional factors governing letter choice.

Table II presents these data in a different form and allows a judgment of the predictive accuracy of each rating category. Tidyman's data are included for comparison. Both the pilot study and Tidyman's data evidence the greater predictive accuracy of Wrong judgments; this rating indicated that a word was very likely to be spelled incorrectly. Doubtful judgments were less reliable predictors of incorrectly spelled words than were Wrong judgments in both studies. Right, however, was a very accurate predictor for Tidyman, while the pilot study data

TABLE II

Percentages and Numbers of Correct and Incorrect Spellings of Words on Which the Judgments Right, Doubtful, and Wrong Were Passed for the Two Procedures Combined and for Tidyman's Study.

Spelling

	Pilot Study						Tidyman (1924)					
	Correct		Incorrect		Total		Correct		Incorrect		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Rating												
Right	30	61%	19	39%	50	37%	8569	93%	675	7%	9244	86%
Doubtful	18	36%	32	64%	48	36%	207	28%	544	72%	751	7%
Wrong	5	14%	30	86%	36	27%	27	5%	545	95%	572	7%
Total	53	---	81	---	134	100%	8803	---	1764	---	10567	100%

indicated it was a fairly unreliable predictor. Thus, either the words to be spelled must be easier words in order to achieve higher rating accuracy with Right judgments, or the S produced spelling must be available to view to achieve accurate predictions of correctly spelled words. In order to gather some notion as to the basis for Ss' ratings, it can be recalled that an analysis of acoustic deficiencies of incorrectly spelled words rated Wrong revealed more frequent deficiencies than those rated Right. Thus words are presumably rated Right when Ss have some confidence in their skill at auditory analysis for spelling purposes. The major difference between subsequently correct words rated Right and those rated Wrong is not primarily due to acoustic deficiencies but rather to incorrect letter choice and misapplication of a structural rule. For those words rated Wrong only a very small percentage were subsequently correctly spelled (in contrast to those rated Right), a fact which is consistent with the notion that this rating category was reserved for those words that presented the S with auditory analysis problems and thus seemed very unlikely to be spelled properly (and were in fact very unlikely to be spelled properly). A χ^2 test of these data indicated that rating and subsequent spelling were significantly related ($\chi^2_2 = 19.24$, $p < .001$).

Table III presents rating accuracy under each of the two procedures. If the distribution of words across the rating categories is compared, it can be noted that Ss rating under Procedure II judged the words as slightly easier to spell than those Ss judging under Procedure I. Additionally, these words were easier for Procedure II Ss who spelled

TABLE III

Percentages and Numbers of Correct and Incorrect Spellings of Words on Which the Judgments Right, Doubtful, and Wrong Were Passed Using Each of the Two Procedures.

Spelling

	Procedure I						Procedure II					
	Correct		Incorrect		Total		Correct		Incorrect		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Rating												
Right	10	59%	7	41%	17	32%	20	63%	12	37%	32	43%
Doubtful	5	21%	19	79%	24	40%	13	50%	13	50%	26	35%
Wrong	0	0%	19	100%	19	28%	5	31%	11	69%	16	22%
Total	15	---	45	---	60	100%	38	---	36	---	74	100%

51% of the words correctly in contrast to Procedure I Ss who spelled only 25% of the words accurately. Thus, any interpretations regarding decision processes used in judging the words as a function of procedure are again confounded with different levels of word difficulty found in the two procedures. Chi-square tests on the data from the two procedures revealed that the relationship between rating and subsequent spelling was much stronger in Procedure I ($\chi^2_2 = 17.38$, $p < .001$) than in Procedure II ($\chi^2_2 = 4.38$, $.10 < p < .20$). When the predictive accuracy of the correct and incorrect categories (excluding Doubtful) is compared for the two procedures, it is clear that Procedure I optimizes accuracy for these two categories, specifically due to the increased predictive accuracy of the Wrong category. Although this increased accuracy might simply be due to the greater word difficulty, it might also be due to the effect of the procedure on Ss' decision making. In Procedure I, Ss heard each word and rated it; they were not required to spell the words until all had been rated. Thus, each word had to be rated exclusively on the basis of the outcome of an auditory analysis; there was no feedback either from viewing the S produced spelling or from difficulties encountered in attempting to produce the spelling that could be used in making subsequent judgments (as were available cues in Procedure II). The predictive accuracy of the Right rating is nearly the same for both procedures, indicating that rating before spelling on the basis of auditory cues alone is not sufficient to make this category a reliable predictor. Adding the opportunity to generate and view spellings of words previous to the target word also does not add to the accuracy of this category as a

predictor of subsequent spelling. What appears to be necessary to an increasingly accurate Right prediction is the opportunity to rate the target word after spelling it, when the visual representation offers information to increase the reliability of this predictor (as per Tidyman, 1924). It must be noted that these conclusions regarding procedure characteristics necessary and sufficient for rating accuracy have not been unequivocally demonstrated, and that future research which removes the present confounding variable must be done to give these notions more than conjectural status.

Protocol Analysis

To provide some data regarding the strategies that Ss use in learning to spell and/or in producing a spelling, Ss were asked about these strategies and their verbal responses were taped. After monitoring these tapes, E formed five categories of strategies and noted which of these were reported at least once by any S. These categories were formed to be as distinct as possible and exhaustive of the verbal responses emitted. The most frequent strategy verbalized was "I sounded out the letters"; all Ss made this response at least once. Three Ss responded that they "broke the word into syllables"; the occurrence of both of these categories indicates sound as a basis for generated spellings; the difference between the strategies is the size of the sound unit forming the basis of the auditory analysis. Five Ss reported some "confusion over letters not sounded (clearly)," for example, Z and S, G and J, able, ness. The protocols of these Ss revealed spelling errors of the options kind. Four Ss alluded to the fact that they

already "knew the word," and two Ss reported "guessing the spelling." What these data indicate is that most Ss are aware of and frequently attempt to use the auditory cues of a word as a basis for producing a spelling and that they additionally are aware that sound cues are not determinate, that these cues are not sufficiently discriminable for accurate letter choice (Z vs. S), or that the same sound can map to several letters.

Summary and Conclusions

The data of the pilot study do not unambiguously support conclusions drawn regarding procedure characteristics which are necessary for higher rating predictability, since procedural comparisons are confounded by the fact that the words were of differing difficulty for Ss in the two procedures. However, because these data and interpretations are consistent with a two-stage process view of spelling, they offer some rational suggestions for the design of instructional programs. These suggestions are herein elaborated with the qualification that future research must be done to substantiate the theoretical models of the spelling process.

For one, it appears that rating accuracy is a function of the features of the stimulus to be rated. Ss can maintain a high predictability for words rated Wrong when ratings are based on sequences of auditory presentations alone; to the extent that other tasks such as writing and spelling are interpolated into the rating of successive auditory inputs, the predictability of both the incorrect rating and the Doubtful rating decreases. The Right rating category was an

unreliable predictor for rating procedures that present the target stimulus orally, regardless of interpolated activity. However, when the S produced spelling is the available information for children when rating, then all three categories become accurate predictors of subsequent spellings.

The argument advanced herein was that in rating auditory inputs, Ss must first judge the sufficiency of the outcome of their auditory analyses for spelling purposes. If these outcomes are judged insufficient, then words are rated Wrong or Doubtful. The reason these categories are such accurate predictors of subsequently incorrect spellings is because accurate auditory analysis is a prerequisite to correct spelling. The categories are more accurate for successive auditory presentations with no other interpolated tasks in that attention to auditory analysis is not disrupted by other tasks. As the auditory analysis is judged likely to be sufficient for spelling, then words are more likely to be predicted Right. Right is an unreliable predictor because, although adequate auditory analysis is necessary for spelling, it is not sufficient. To make this category a more reliable predictor, cues which determine choice of letters must be salient and available for Ss to use in making rating judgments. These cues (such as position of the sound in a syllable and adjacent phonemes) are strongly available (although perhaps not salient) when S produced spellings are read and form the basis for rating.

These data make some suggestions for the design of CAI drill-and-practice spelling programs. Although these data do not point to a

unique instructional design, one design which is consistent with them is the following: First, if ratings are to be used for prescribing instruction, then ratings should be used when they are accurate predictors and in situations where they can offer diagnostic information. When ratings are made on auditory stimuli, then judgments of Wrong and Doubtful presumably identify those words with which S at least has difficulty with an auditory analysis. With words rated Right, acoustic problems are probably less likely to be sources of spelling errors. Since it is very difficult to separate misspellings arising primarily from auditory problems from those arising from option choice, then it is reasonable to let S differentiate these. If all words are presented auditorily and each word is rated in turn (with no interpolated tasks), then the subset of words rated Doubtful and Wrong can be designated for diagnosis and treatment in an auditory training sequence in addition to options training. For words rated "correct," S would be asked to spell this subset, and the S produced spellings would be placed in the same order for re-presentation and also rating (so that auditory memory for each word would decay or be interfered with as time passes or new items are presented between hearing the word and rating the spelling, so that S must read his spelling). Words rated Right and spelled correctly would be given no further treatment; for those words in which rating and spelling do not agree, these words would be given instructional treatment in the context of the optional rules which S has not mastered. Those rules would be chosen by an analysis of the incorrect option choices in the case of

an incorrect spelling, and by a determination of the optional rules which were most likely to be less well learned in the case of a correct spelling. The latter decision would probably be based on spelling literature which provides information regarding the differential difficulties of various spelling generalizations.

References

Banks, W. P. Signal detection theory and human memory. Psychological Bulletin, 1970, 74, 81-99.

Bernbach, H. A. Decision processes in memory. Psychological Review, 1967, 74, 462-480.

Block, K. K. & Butler, P. A. Curriculum development in CAI spelling: Psychological considerations and instructional components. In preparation, Learning Research and Development Center, University of Pittsburgh, 1971.

Bower, G. H. & Winzenz, D. Group structure, coding, and memory for digit series. Journal of Experimental Psychology Monographs, 1969, 80, No. 2, Part 2, 1-17.

Diggory, J. C. Responses to experimentally induced failure. American Journal of Psychology, 1949, 62, 48-61.

Donaldson, W. & Glathe, H. Signal detection analysis of recall and recognition memory. Canadian Journal of Psychology, 1970, 24, 42-56.

Hendrickson, G. & Pechstein, L. A. The spelling consciousness of college students. Journal of Educational Psychology, 1926, 17, 37-44.

Lewin, K., Dembo, T., Festinger, L., & Sears, P. Level of aspiration. In J. McV. Hunt (Ed.), Personality and the behavior disorders, Vol. I. New York: Ronald Press, 1944. Pp. 333-378.

Murdock, B. B., Jr. The criterion problem in short-term memory. Journal of Experimental Psychology, 1966, 72, 317-324.

Restle, F. Sources of difficulty in learning paired associates. In R. C. Atkinson (Ed.), Studies in mathematical psychology. Stanford: Stanford University Press, 1964. Pp. 116-172.

Ricciuti, H. N. A review of procedural variations in level of aspiration studies. Lackland, A. F. Base, Texas: U. S. A. F. Human Resources Research Center, Research Bulletin, 1951, 24-51.

Tidyman, W. F. Do elementary school pupils know when they make mistakes in spelling? School and Society, 1924, 20, 349-350.

Trabasso, T. & Bower, G. H. Attention in learning: Theory and research. New York: John Wiley and Sons, Inc., 1968.

Wickelgren, W. A. & Norman, D. A. Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology, 1966, 3, 316-347.